



09/914733
PCT/AU00/00158

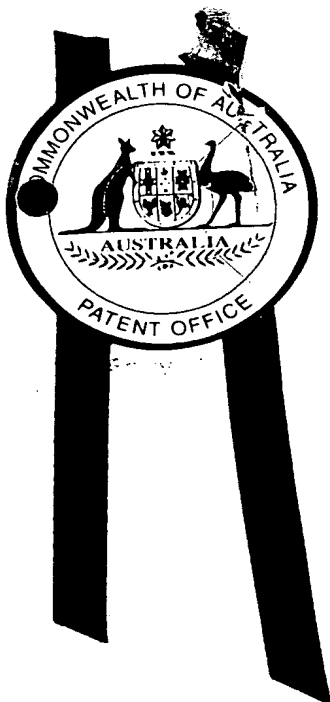
REC'D 18 APR 2000	
WIPO	PCT

AU 00/158
4

Patent Office
Canberra

I, ANNA MAIJA MADL, ACTING TEAM LEADER EXAMINATION SUPPORT & SALES hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. PP 9048 for a patent by CLAIRVIEW INTERNET PTY LTD filed on 04 March 1999.

I further certify that the name of the applicant has been amended to TEL.NET MEDIA PTY LTD pursuant to the provisions of Section 104 of the Patents Act 1990.



WITNESS my hand this
Sixteenth day of March 2000

A. M. Madl

ANNA MAIJA MADL
ACTING TEAM LEADER
EXAMINATION SUPPORT & SALES

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)



Tel. Net Media Pty Ltd
~~CLAIRVIEW INTERNET PTY LTD~~

FORM 9

COMMONWEALTH OF AUSTRALIA

Patents Act 1990

PROVISIONAL SPECIFICATION FOR THE INVENTION ENTITLED:

"APPARATUS AND SYSTEM FOR CLASSIFYING AND CONTROL
ACCESS TO INFORMATION"

This invention is described in the following statement:

Apparatus and system for classifying and control access to information

TECHNICAL FIELD OF THE INVENTION

THIS INVENTION relates to apparatus and system for classifying information on communications network and in particular but not limited to apparatus and system for classifying content servers and for selectively controlling access to classified content servers.

BACKGROUND OF THE INVENTION

The phenomenon growth of information technology has allowed many people to have access to diverse information on communications networks. The Internet in particular allows fetching of information from any cooperating computers or content servers located in different parts of the world by simply clicking references to the information. As the number of accessible computers or content servers and the amount of information over the communications network grow daily it becomes increasing difficult to classify them manually.

Known systems for controlling the types of information accessible on a network rely on comparing a requested destination with those on pre-determined Access Control Lists (ACL) or on word matching to determine whether to allow or deny access. This approach can be applied at the client node prior to requesting the information or on any suitably intelligent network device capable of intercepting the request or subsequent reply prior to it reaching the requester. For example, in the case of an Internet browser running on a PC or work station, a request is made for an Internet resource such as a web site. Software monitoring such requests on the PC can be configured to scan a pre-determined list of site addresses for a match. If found, access to the site may be denied and a suitable

message displayed informing the user. Alternatively, the request may be allowed to proceed, but as data is received from the site it is scanned for any of a set of pre-determined words, word fragments or phrases. If a match is found the site is not displayed but instead replaced with a suitable message. Typically, this type of control software is installed on a PC or work station which does not have particularly strict access privileges. The control software can be easily removed, disabled or otherwise circumvented and thereby defeating the control system.

A network device capable of intercepting the request or reply, such as a proxy server, may perform similar actions using the same methods of web site matching. This is usually maintained by a network administrator with strict access rights. Also, any clients that must pass through the device in order to access the network can have content control enforced. This allows content control of multiple clients from one central point.

While these known systems do provide some access control abilities, there are several disadvantages. A system based on word or phrase matching can only match text and it therefore would allow access to undesired information comprising graphic images. Also, a single word may match a broad range of sites with quite different classes of information. As an example, when the word "sex" is used to match pornographic sites the system would also block access to other sites providing non offensive information such as articles on biology.

A system based on access control lists is much more selective. Access would only be denied to sites contained in the lists. While a suitably large list could bar access to a great deal of undesirable information it is difficult to keep

up to date due to the rapid increase in the number of new sites and removal of sites.

The above systems also do not lend themselves to adaptation to other network protocols and services such as interactive chat, streaming video, email or encrypted data streams. Extending to different languages also poses a problem for globalisation of these systems.

OBJECT OF THE INVENTION

An object of the present invention is to alleviate or to reduce to a certain degree one or more of the above disadvantages.

Another object of the present invention is provide an apparatus/system for classifying user profiles.

SUMMARY OF THE INVENTION

In one aspect therefor the present invention resides in an apparatus for classifying information on communications network. The apparatus comprises means for obtaining one or more transmission characteristics of information on a path of said communications network, analysing means for predicting a classification of said information based on said one or more transmission characteristics.

In a second aspect therefor the present invention resides in an apparatus for classifying content servers accessible on communications network. The apparatus comprises means for obtaining one or more transmission characteristics of information provided by any of said content servers on a path of said communications network, analysing means for predicting a classification of said information based on said one or more transmission characteristics.

In a third aspect therefor the present invention resides in a computer program for classifying information on communications network. The program comprises means for obtaining one or more transmission characteristics of information on a path of said communications network, analysing means for
5 predicting a classification of said information based on said one or more transmission characteristics.

In a fourth aspect therefor the present invention resides in a computer program for classifying content servers accessible on communications network. The apparatus comprises means for obtaining one or more transmission
10 characteristics of information provided by any of said content servers on a path of said communications network, analysing means for predicting a classification of said information based on said one or more transmission characteristics.

In a fifth aspect therefor the present invention resides in an apparatus/computer program for classifying user profiles of users accessing
15 information or content terminals on communications network. The apparatus/computer program comprises means for obtaining one or more transmission characteristics of information or information provided by any of said content servers, on a path of said communications network, analysing means for predicting a classification of said information or content server based on said one
20 or more transmission characteristics, and means for classifying user profile in accordance with the predicted classification.

The above invention may also comprise means for storing said one or more transmission characteristics.

In order that the present invention can be more readily understood and be put into practical effect reference will now be made to the accompanying drawings which illustrate one preferred embodiment of the invention and wherein:

BRIEF DESCRIPTION OF THE DRAWING

5 Figure 1 is a schematic diagram of the apparatus according to the invention;

Figure 2 is a table of selected data of captured packets of a search engine using the apparatus shown in Figure 1;

10 Figure 3 is a partial table of selected data of captured packets of a news web site using the apparatus shown in Figure 1;

Figure 4 is a table of selected data of captured packets of an entertainment web site using the apparatus shown in Figure 1;

Figure 5 is a table of selected data of captured packets of the web site of an e-commerce merchant using the apparatus shown in Figure 1;

15 Figure 6 is a table of selected data of captured packets of the web site of another e-commerce merchant using the apparatus shown in Figure 1;

Figure 7 is a table of selected data of captured packets of a pornography web site using the apparatus shown in Figure 1;

20 Figure 8 is a table of selected data of captured packets of another pornography web site using the apparatus shown in Figure 1;

Figure 9 is a table of model N1 results using the apparatus shown in Figure 1;

Figure 10 is a table of model N2 results using the apparatus shown in Figure 1;

Figure 11 is a table of model N3 results using the apparatus shown in Figure 1; and

Figure 12 is a table of classification prediction confidence levels using the apparatus shown in Figure 1.

5

DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring initially to Figure 1 there is shown an apparatus 10 for classifying media or information flowing a path of a communications network which in this case is the Internet.

As can be seen, network traffic passing through the apparatus 10 is captured and analysed for providing statistics relating to interactions between two or more terminals (not shown). The captured traffic is first checked against a list of predetermined classifications to determine if it is known or unknown.

When the captured traffic is of an unknown classification, various models (to be described more fully below) are applied to the data set in the captured traffic in order to predict the content classification. The models use parameters derived from a knowledge base of previously classified data sets. The model of best fit determines the classification of the content of the newly captured traffic. Thus, the web site sending the captured traffic is now classified and is added to the list of known classifications.

20

Following classification the captured data set is stored in the knowledge base. As the knowledge base expands, more data are used for the model parameters. This refines the apparatus and results in improved predictive performance.

The sites that are deemed are added to Access control lists (ACLs). ACLs are used control the flow of content between terminals. E.g. Undesired content can be prevented from travelling further through the network by simply not forwarding it, or by replacing it, or by intercepting the request for such content and modifying its destination.

Classification of traffic from content servers are relatively static. On the other hand, user terminals that interact with these content servers are variable and their classifications are considered transient classifications.

Whereas classifications of content servers form a model of the style of content residing on the server, transient classifications form a model of style of content being viewed by a user terminal, or content consumer. This in effect forms a behaviour profile of such a consumer. This profile can be used to tailor the content to suit the consumer.

As mentioned earlier the apparatus 10 captures a set of observed data relating to a network interaction event, and provides a set of results indicating the classification of a resource or personality residing at each network node involved in the interaction. This is accomplished by applying various statistical models to a profile, and testing this against results obtained from profiles of known classifications. In this example of the invention this process is represented by the

following formulas:

x is an unknown profile to be classified;

Profiles $p_1, p_2, p_3 \dots p_n$ are of known classifications;

Models $M_1, M_2, M_3 \dots M_n$ are available to operate on these profiles; and

$C_1, C_2, C_3 \dots C_n$ are profile classifications.

The population of a profile of classification C1, may be defined by the population

of M1(p). M1(x) may be tested against the true population using standard statistical hypothesis methods.

- 5 A pre-determined set of media terminals of a classification are modelled by various models M1, M2 .. Mn. Each model consists of an approach and a set of parameter, e.g linear regression, gradient and point of interception, so that for a single classification M1(p1,p2 .. pn), M2(q1,q2 .. qn) .. Mn(r1,r2 .. rn) are used to model the population from the classification. The models may be based on
- 10 mathematical structures, or arbitrary rules.

The models are continually refined as more network traffic passes through the APPARATUS, thereby increasing the population space from which the classifications are computed.

A terminal may be permanently or transitionally defined to a classification.

- 15 A transitionally defined terminal may move between classifications based on the fit of the observed traffic to the models of the various classifications.

- Figures 2 to 8 are tables of selected data of traffic for to testing the profile of network interaction with a content server to determine if it contains media content of a pornographic nature. Assumption is made that profiles for content
- 20 servers contain a variable which is the average size of graphical images served.

A normal distribution or similar non-deterministic probability distribution is then used to test the hypothesis that the profile belongs to a population classified as pornographic. In this example, the population of the classification may be defined by the population of N(a,b) where N is the image size and a and

b are the mean and variance respectively, based on a normal distribution. The average and standard deviation derived from the observed samples is tested against the true population using standard statistical hypothesis methods.

In some cases this approach may be broadened to encompass analysis of variance methods with multiple dependant variables, to model the characteristics of a site. Traditional ANOVA may be applied to model the media content.

A variety of traditional deterministic and non-deterministic models may be applied to determine the hypothesis of profile classification. These may be changed or upgraded continually depending on the level of predictive power found. The functionality of models used is not limited to, but can include simple rules-of-thumb, deterministic and non-deterministic probability models, or arbitrary calculations.

The choice of model is primarily dictated by the predictive power of that model against the population in question.

Figures 2 through 8 show examples of basic data set that can be gathered by observing network traffic of a typical interaction between a client browser and a web server.

Figures 9 to 11 illustrate a simple classification model. This model looks at the size, content and relationships of objects being transmitted by a content server. The outcome of this model is to determine if the media being transmitted has pornographic content.

Classification: pornographic

Standard Model:

N1(a,b)

Where $N1$ is the image size, a and b are the mean and variance respectively, based on a normal distribution.

$N2(c,d)$

Where $N2$ is the ratio of text to graphics, c and d are the total size of the text and graphic objects respectively.

$N3(e)$

Where $N3$ is the count of word patterns matched from a list of pre-determined words, and e is the text of an object.

Observed Samples are given in the tables shown in Figures 2 to 8.

For model $N1$ shown in Figure 9, there is applied the normal distribution hypothesis test to the observed samples deriving the results.

The result shows confidence to the 93% and 87% level for sites 6 and 7 respectively, that the sites belong to a population of pornographic sites. The other samples give much lower confidence levels.

For model $N2$ shown in Figure 10, a simple rule is used to test if the ratio is below a pre-determined threshold. The results show that sites 2, 4, 6 and 7 are within the threshold rating.

For Model $N3$ shown in Figure 11, a simple rule is used to test if the number of words matching a list of patterns, exceeds a pre-determined threshold.

The results show that sites 6 and 7 exceed the threshold.

A weighting formula is then applied to derive a final result as shown in Figure 12.

Therefore, using this example model, the apparatus 10 would predict that sites 6 and 7 are probably serving media with pornographic content, whereas sites 1 through 5 probably are not.

5 Whilst the above has been given by way of illustrative example of the present invention many variations and modifications thereto will be apparent to those skilled in the art without departing from the broad ambit and scope of the invention as herein set forth.

DATED this 4th day of March 1999

Tel. Net Media Pty Ltd
~~CLAIRVIEW PTY LTD~~
By their Patent Attorneys
INTELLPRO



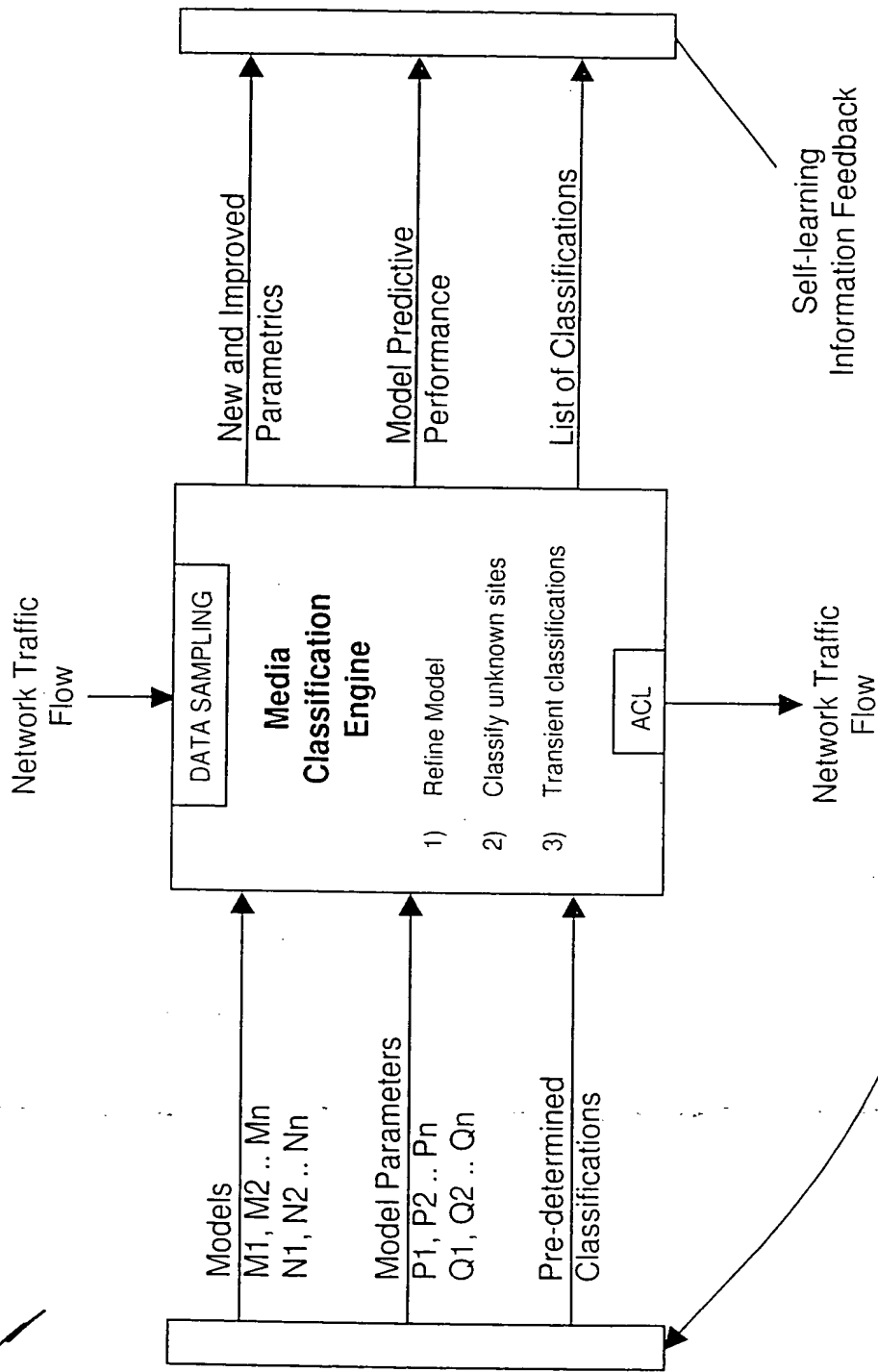


FIG. 1

Site 1 Objects - Content Type "Search Engine"							
Object	Source		Destination		Type	Size	Timestamp
	IP Address	Port	IP Address	Port			
1	202.139.16.45	63450	204.71.200.72	80	GET	36	14:53:17
2	204.71.200.72	80	202.139.16.45	63450	text/html	9424	14:53:17
3	202.139.16.45	63450	204.71.200.72	80	GET	79	14:53:19
4	202.139.16.45	63450	204.71.200.72	80	GET	46	14:53:19
5	202.139.16.45	63450	204.71.200.72	80	GET	50	14:53:19
6	204.71.200.72	80	202.139.16.45	63450	image/gif	2637	14:53:19
7	204.71.200.72	80	202.139.16.45	63450	image/gif	4672	14:53:20
8	204.71.200.72	80	202.139.16.45	63450	image/gif	357	14:53:20
9	202.139.16.45	63450	204.71.200.72	80	GET	56	14:53:27
10	204.71.200.72	80	202.139.16.45	63450	text/html	11193	14:53:28
11	202.139.16.45	63450	204.71.200.72	80	GET	59	14:53:29
12	204.71.200.72	80	202.139.16.45	63450	image/gif	11522	14:53:29
13	202.139.16.45	63450	204.71.200.72	80	GET	67	14:53:30
14	202.139.16.45	63450	204.71.200.72	80	GET	67	14:53:30
15	202.139.16.45	63450	204.71.200.72	80	GET	59	14:53:30
16	204.71.200.72	80	202.139.16.45	63450	image/gif	1398	14:53:30
17	202.139.16.45	63450	204.71.200.72	80	GET	76	14:53:30
18	204.71.200.72	80	202.139.16.45	63450	image/gif	1728	14:53:30
19	202.139.16.45	63450	204.71.200.72	80	GET	69	14:53:30
20	204.71.200.72	80	202.139.16.45	63450	image/gif	962	14:53:30
21	204.71.200.72	80	202.139.16.45	63450	image/gif	946	14:53:31
22	204.71.200.72	80	202.139.16.45	63450	image/gif	1716	14:53:31

FIG. 2

Site 2 Objects - Content Type "News"							
Object	Source		Destination		Type	Size	Timestamp
	IP Address	Port	IP Address	Port			
1	202.139.16.45	63450	165.69.1.187	80	GET	38	14:54:04
2	165.69.1.187	80	202.139.16.45	63450	text/html	2312	14:54:04
3	202.139.16.45	63450	165.69.1.187	80	GET	52	14:54:05
4	202.139.16.45	63450	165.69.1.187	80	GET	56	14:54:05
5	202.139.16.45	63450	165.69.1.187	80	GET	46	14:54:05
6	202.139.16.45	63450	165.69.1.187	80	GET	48	14:54:05
7	202.139.16.45	63450	165.69.1.187	80	GET	47	14:54:05
8	202.139.16.45	63450	165.69.1.187	80	GET	47	14:54:05
9	165.69.1.187	80	202.139.16.45	63450	text/html	333	14:54:05
10	165.69.1.187	80	202.139.16.45	63450	text/html	56	14:54:06
11	165.69.1.187	80	202.139.16.45	63450	text/html	2445	14:54:06
12	165.69.1.187	80	202.139.16.45	63450	text/html	202	14:54:06
13	165.69.1.187	80	202.139.16.45	63450	text/html	202	14:54:06
14	165.69.1.187	80	202.139.16.45	63450	text/html	56	14:54:06
15	202.139.16.45	63450	165.69.1.187	80	GET	71	14:54:06
16	165.69.1.187	80	202.139.16.45	63450	image/gif	1229	14:54:06
17	202.139.16.45	63450	165.69.1.187	80	GET	64	14:54:06
18	165.69.1.187	80	202.139.16.45	63450	image/gif	43	14:54:06
19	202.139.16.45	63450	165.69.1.187	80	GET	64	14:54:06
20	202.139.16.45	63450	165.69.1.187	80	GET	64	14:54:07
21	165.69.1.187	80	202.139.16.45	63450	image/gif	43	14:54:07
22	165.69.1.187	80	202.139.16.45	63450	image/gif	43	14:54:07
23	202.139.16.45	63450	165.69.1.187	80	GET	67	14:54:07
24	165.69.1.187	80	202.139.16.45	63450	image/gif	2442	14:54:07
25	202.139.16.45	63450	165.69.1.187	80	GET	73	14:54:07
26	165.69.1.187	80	202.139.16.45	63450	image/gif	1364	14:54:07
27	202.139.16.45	63450	165.69.1.187	80	GET	71	14:54:07
28	165.69.1.187	80	202.139.16.45	63450	image/gif	8942	14:54:07
29	202.139.16.45	63450	165.69.1.187	80	GET	71	14:54:07
30	202.139.16.45	63450	165.69.1.187	80	GET	65	14:54:08
31	165.69.1.187	80	202.139.16.45	63450	unknown	10	14:54:08
32	165.69.1.187	80	202.139.16.45	63450	image/gif	15550	14:54:08
33	202.139.16.45	63450	165.69.1.187	80	GET	77	14:54:08
34	165.69.1.187	80	202.139.16.45	63450	image/gif	4732	14:54:08
35	202.139.16.45	63450	165.69.1.187	80	GET	70	14:54:08
36	202.139.16.45	63450	165.69.1.187	80	GET	70	14:54:09
37	202.139.16.45	63450	165.69.1.187	80	GET	68	14:54:09
38	202.139.16.45	63450	165.69.1.187	80	GET	68	14:54:09
39	165.69.1.187	80	202.139.16.45	63450	image/gif	436	14:54:09
40	165.69.1.187	80	202.139.16.45	63450	image/gif	405	14:54:09
41	165.69.1.187	80	202.139.16.45	63450	image/gif	436	14:54:09
42	165.69.1.187	80	202.139.16.45	63450	image/gif	405	14:54:09

FIG. 3

Site 3 Objects - Content Type "Entertainment"							
Object	Source		Destination		Type	Size	Timestamp
	IP Address	Port	IP Address	Port			
1	202.139.16.45	63450	204.202.129.23	80	GET	40	14:55:10
2	204.202.129.23	80	202.139.16.45	63450	text/html	56	14:55:11
3	202.139.16.45	63450	204.202.129.23	80	GET	39	14:55:11
4	204.202.129.23	80	202.139.16.45	63450	text/html	25765	14:55:12
5	202.139.16.45	63450	204.202.129.23	80	GET	58	14:55:13
6	202.139.16.45	63450	204.202.129.23	80	GET	66	14:55:13
7	202.139.16.45	63450	204.202.129.23	80	GET	62	14:55:13
8	204.202.129.23	80	202.139.16.45	63450	image/jpeg	5277	14:55:14
9	204.202.129.23	80	202.139.16.45	63450	image/gif	43	14:55:14
10	204.202.129.23	80	202.139.16.45	63450	image/gif	1266	14:55:14
11	202.139.16.45	63450	204.202.129.23	80	GET	63	14:55:20
12	202.139.16.45	63450	204.202.129.23	80	GET	58	14:55:21
13	202.139.16.45	63450	204.202.129.23	80	GET	72	14:55:21
14	204.202.129.23	80	202.139.16.45	63450	image/gif	1733	14:55:22
15	204.202.129.23	80	202.139.16.45	63450	image/gif	5314	14:55:22
16	204.202.129.23	80	202.139.16.45	63450	image/gif	414	14:55:22
17	202.139.16.45	63450	204.202.129.23	80	GET	68	14:55:22
18	202.139.16.45	63450	204.202.129.23	80	GET	62	14:55:22
19	204.202.129.23	80	202.139.16.45	63450	image/gif	406	14:55:22
20	204.202.129.23	80	202.139.16.45	63450	image/gif	746	14:55:22
21	202.139.16.45	63450	204.202.129.23	80	GET	65	14:55:23
22	202.139.16.45	63450	204.202.129.23	80	GET	58	14:55:23
23	202.139.16.45	63450	204.202.129.23	80	GET	63	14:55:23
24	202.139.16.45	63450	204.202.129.23	80	GET	62	14:55:23
25	204.202.129.23	80	202.139.16.45	63450	image/gif	1665	14:55:23
26	204.202.129.23	80	202.139.16.45	63450	image/gif	35	14:55:24
27	204.202.129.23	80	202.139.16.45	63450	image/gif	906	14:55:24
28	204.202.129.23	80	202.139.16.45	63450	image/gif	447	14:55:24
29	202.139.16.45	63450	204.202.129.23	80	GET	67	14:55:24
30	202.139.16.45	63450	204.202.129.23	80	GET	58	14:55:24
31	202.139.16.45	63450	204.202.129.23	80	GET	62	14:55:24
32	204.202.129.23	80	202.139.16.45	63450	image/jpeg	7861	14:55:24
33	204.202.129.23	80	202.139.16.45	63450	image/gif	391	14:55:25
34	204.202.129.23	80	202.139.16.45	63450	image/gif	641	14:55:25
35	202.139.16.45	63450	204.202.129.23	80	GET	57	14:55:25
36	204.202.129.23	80	202.139.16.45	63450	image/gif	377	14:55:25
37	202.139.16.45	63450	204.202.129.23	80	GET	60	14:55:26
38	202.139.16.45	63450	204.202.129.23	80	GET	78	14:55:26
39	202.139.16.45	63450	204.202.129.23	80	GET	74	14:55:26
40	202.139.16.45	63450	204.202.129.23	80	GET	67	14:55:26
41	204.202.129.23	80	202.139.16.45	63450	image/gif	403	14:55:26
42	204.202.129.23	80	202.139.16.45	63450	image/gif	1796	14:55:26
43	204.202.129.23	80	202.139.16.45	63450	image/gif	6845	14:55:27
44	202.139.16.45	63450	204.202.129.23	80	GET	56	14:55:27
45	204.202.129.23	80	202.139.16.45	63450	image/jpeg	17796	14:55:27
46	202.139.16.45	63450	204.202.129.23	80	GET	56	14:55:27
47	204.202.129.23	80	202.139.16.45	63450	image/gif	49	14:55:27
48	204.202.129.23	80	202.139.16.45	63450	image/gif	44	14:55:27

FIG. 4

Site 4 Objects - Content Type "Computer Hardware"							
Object	Source		Destination		Type	Size	Timestamp
	IP Address	Port	IP Address	Port			
1	202.139.16.45	63450	209.143.240.6	80	GET	34	14:55:49
2	209.143.240.6	80	202.139.16.45	63450	text/html	1645	14:55:50
3	202.139.16.45	63450	209.143.240.6	80	GET	47	14:55:51
4	209.143.240.6	80	202.139.16.45	63450	text/html	56	14:55:51
5	202.139.16.45	63450	209.143.240.6	80	GET	64	14:55:53
6	202.139.16.45	63450	209.143.240.6	80	GET	63	14:55:53
7	202.139.16.45	63450	209.143.240.6	80	GET	63	14:55:53
8	209.143.240.6	80	202.139.16.45	63450	image/gif	290	14:55:53
9	202.139.16.45	63450	209.143.240.6	80	GET	64	14:55:54
10	209.143.240.6	80	202.139.16.45	63450	image/gif	403	14:55:54
11	209.143.240.6	80	202.139.16.45	63450	image/gif	381	14:55:54
12	202.139.16.45	63450	209.143.240.6	80	GET	63	14:55:54
13	202.139.16.45	63450	209.143.240.6	80	GET	65	14:55:54
14	209.143.240.6	80	202.139.16.45	63450	image/gif	348	14:55:54
15	202.139.16.45	63450	209.143.240.6	80	GET	65	14:55:54
16	202.139.16.45	63450	209.143.240.6	80	GET	65	14:55:54
17	209.143.240.6	80	202.139.16.45	63450	image/gif	354	14:55:54
18	209.143.240.6	80	202.139.16.45	63450	image/gif	600	14:55:54
19	202.139.16.45	63450	209.143.240.6	80	GET	65	14:55:55
20	202.139.16.45	63450	209.143.240.6	80	GET	54	14:55:55
21	209.143.240.6	80	202.139.16.45	63450	image/gif	490	14:55:55
22	202.139.16.45	63450	209.143.240.6	80	GET	53	14:55:55
23	209.143.240.6	80	202.139.16.45	63450	image/gif	571	14:55:55
24	209.143.240.6	80	202.139.16.45	63450	image/gif	322	14:55:55
25	209.143.240.6	80	202.139.16.45	63450	image/gif	571	14:55:55
26	202.139.16.45	63450	209.143.240.6	80	GET	55	14:55:55
27	209.143.240.6	80	202.139.16.45	63450	image/gif	363	14:55:55
28	202.139.16.45	63450	209.143.240.6	80	GET	63	14:55:55
29	202.139.16.45	63450	209.143.240.6	80	GET	63	14:55:55
30	202.139.16.45	63450	209.143.240.6	80	GET	62	14:55:55
31	209.143.240.6	80	202.139.16.45	63450	image/gif	241	14:55:56
32	202.139.16.45	63450	209.143.240.6	80	GET	59	14:55:56
33	209.143.240.6	80	202.139.16.45	63450	image/gif	488	14:55:56
34	209.143.240.6	80	202.139.16.45	63450	image/gif	463	14:55:56
35	202.139.16.45	63450	209.143.240.6	80	GET	59	14:55:56
36	209.143.240.6	80	202.139.16.45	63450	image/gif	714	14:55:56
37	202.139.16.45	63450	209.143.240.6	80	GET	53	14:55:56
38	209.143.240.6	80	202.139.16.45	63450	image/gif	35	14:55:56
39	202.139.16.45	63450	209.143.240.6	80	GET	53	14:55:56
40	209.143.240.6	80	202.139.16.45	63450	image/gif	1188	14:55:56
41	202.139.16.45	63450	209.143.240.6	80	GET	54	14:55:56
42	209.143.240.6	80	202.139.16.45	63450	image/gif	327	14:55:57

FIG. 5

Site 5 Objects - Content Type "Computer Software"							
Object	Source		Destination		Type	Size	Timestamp
	IP Address	Port	IP Address	Port			
1	202.139.16.45	63450	207.46.130.149	80	GET	40	14:57:46
2	207.46.130.149	80	202.139.16.45	63450	text/html	19019	14:57:47
3	202.139.16.45	63450	207.46.130.149	80	GET	66	14:57:48
4	207.46.130.149	80	202.139.16.45	63450	application	18768	14:57:49
5	202.139.16.45	63450	207.46.130.149	80	GET	71	14:57:50
6	207.46.130.149	80	202.139.16.45	63450	application	2941	14:57:50
7	202.139.16.45	63450	207.46.130.149	80	GET	70	14:57:50
8	207.46.130.149	80	202.139.16.45	63450	application	2812	14:57:51
9	202.139.16.45	63450	207.46.130.149	80	GET	75	14:57:51
10	202.139.16.45	63450	207.46.130.149	80	GET	72	14:57:51
11	202.139.16.45	63450	207.46.130.149	80	GET	80	14:57:51
12	202.139.16.45	63450	207.46.130.149	80	GET	73	14:57:51
13	207.46.130.149	80	202.139.16.45	63450	image/gif	7976	14:57:51
14	202.139.16.45	63450	207.46.130.149	80	GET	80	14:57:52
15	207.46.130.149	80	202.139.16.45	63450	image/gif	139	14:57:52
16	207.46.130.149	80	202.139.16.45	63450	image/gif	44	14:57:52
17	207.46.130.149	80	202.139.16.45	63450	image/gif	666	14:57:52
18	202.139.16.45	63450	207.46.130.149	80	GET	80	14:57:52
19	207.46.130.149	80	202.139.16.45	63450	image/gif	44	14:57:52
20	207.46.130.149	80	202.139.16.45	63450	image/gif	44	14:57:52
21	202.139.16.45	63450	207.46.130.149	80	GET	80	14:57:52
22	207.46.130.149	80	202.139.16.45	63450	image/gif	44	14:57:52
23	202.139.16.45	63450	207.46.130.149	80	GET	74	14:57:52
24	202.139.16.45	63450	207.46.130.149	80	GET	78	14:57:53
25	202.139.16.45	63450	207.46.130.149	80	GET	83	14:57:53
26	202.139.16.45	63450	207.46.130.149	80	GET	77	14:57:53
27	207.46.130.149	80	202.139.16.45	63450	image/gif	1348	14:57:53
28	207.46.130.149	80	202.139.16.45	63450	image/gif	54	14:57:53
29	207.46.130.149	80	202.139.16.45	63450	image/gif	3966	14:57:53
30	207.46.130.149	80	202.139.16.45	63450	image/gif	1883	14:57:53
31	202.139.16.45	63450	207.46.130.149	80	GET	77	14:57:53
32	207.46.130.149	80	202.139.16.45	63450	image/gif	1044	14:57:54

FIG. 6

Site 6 Objects - Content Type "Pornography"							
Object	Source		Destination		Type	Size	Timestamp
	IP Address	Port	IP Address	Port			
1	202.139.16.45	63450	207.87.4.205	80	GET	36	12:01:53
2	207.87.4.205	80	202.139.16.45	63450	text/html	4254	12:01:54
3	202.139.16.45	63450	207.87.4.205	80	GET	46	12:01:56
4	207.87.4.205	80	202.139.16.45	63450	image/jpeg	4029	12:01:56
5	202.139.16.45	63450	207.87.4.205	80	GET	46	12:02:00
6	207.87.4.205	80	202.139.16.45	63450	text/html	5615	12:02:01
7	202.139.16.45	63450	207.87.4.205	80	GET	48	12:02:01
8	202.139.16.45	63450	207.87.4.205	80	GET	48	12:02:01
9	202.139.16.45	63450	207.87.4.205	80	GET	44	12:02:01
10	202.139.16.45	63450	207.87.4.205	80	GET	48	12:02:01
11	207.87.4.205	80	202.139.16.45	63450	image/gif	7026	12:02:01
12	202.139.16.45	63450	207.87.4.205	80	GET	45	12:02:02
13	207.87.4.205	80	202.139.16.45	63450	image/jpeg	20063	12:02:02
14	207.87.4.205	80	202.139.16.45	63450	image/jpeg	20526	12:02:02
15	207.87.4.205	80	202.139.16.45	63450	image/jpeg	31751	12:02:02
16	207.87.4.205	80	202.139.16.45	63450	image/gif	723	12:02:02
17	202.139.16.45	63450	207.87.4.205	80	GET	46	12:02:02
18	207.87.4.205	80	202.139.16.45	63450	image/gif	4399	12:02:03
19	202.139.16.45	63450	207.87.4.205	80	GET	46	12:02:03
20	202.139.16.45	63450	207.87.4.205	80	GET	46	12:02:04
21	207.87.4.205	80	202.139.16.45	63450	image/gif	8928	12:02:04
22	207.87.4.205	80	202.139.16.45	63450	image/gif	5474	12:02:04

FIG. 7

Site 7 Objects - Content Type "Pornography"							
Object	Source		Destination		Type	Size	Timestamp
	IP Address	Port	IP Address	Port			
1	202.139.16.45	63450	209.164.26.157	80	GET	38	12:02:15
2	209.164.26.157	80	202.139.16.45	63450	text/html	301	12:02:16
3	202.139.16.45	63450	209.164.26.157	80	GET	49	12:02:16
4	209.164.26.157	80	202.139.16.45	63450	text/html	4211	12:02:16
5	202.139.16.45	63450	209.164.26.157	80	GET	56	12:02:17
6	209.164.26.157	80	202.139.16.45	63450	image/gif	290839	12:02:17
7	202.139.16.45	63450	209.164.26.157	80	GET	64	12:02:21
8	209.164.26.157	80	202.139.16.45	63450	text/html	12066	12:02:21
9	202.139.16.45	63450	209.164.26.157	80	GET	76	12:02:22
10	209.164.26.157	80	202.139.16.45	63450	image/gif	4675	12:02:22
11	202.139.16.45	63450	209.164.26.157	80	GET	76	12:02:22
12	202.139.16.45	63450	209.164.26.157	80	GET	76	12:02:22
13	209.164.26.157	80	202.139.16.45	63450	image/gif	4675	12:02:22
14	209.164.26.157	80	202.139.16.45	63450	image/gif	4822	12:02:22
15	202.139.16.45	63450	209.164.26.157	80	GET	76	12:02:23
16	209.164.26.157	80	202.139.16.45	63450	image/gif	4812	12:02:23
17	202.139.16.45	63450	209.164.26.157	80	GET	76	12:02:23
18	209.164.26.157	80	202.139.16.45	63450	image/gif	4824	12:02:24
19	202.139.16.45	63450	209.164.26.157	80	GET	76	12:02:24
20	209.164.26.157	80	202.139.16.45	63450	image/gif	4858	12:02:24
21	202.139.16.45	63450	209.164.26.157	80	GET	76	12:02:24
22	209.164.26.157	80	202.139.16.45	63450	image/gif	4804	12:02:24
23	202.139.16.45	63450	209.164.26.157	80	GET	76	12:02:24
24	209.164.26.157	80	202.139.16.45	63450	image/gif	4800	12:02:24
25	202.139.16.45	63450	209.164.26.157	80	GET	76	12:02:24
26	209.164.26.157	80	202.139.16.45	63450	image/gif	4767	12:02:25
27	202.139.16.45	63450	209.164.26.157	80	GET	76	12:02:25
28	209.164.26.157	80	202.139.16.45	63450	image/gif	4732	12:02:25
29	202.139.16.45	63450	209.164.26.157	80	GET	159	12:02:25
30	202.139.16.45	63450	209.164.26.157	80	GET	159	12:02:25
31	202.139.16.45	63450	209.164.26.157	80	GET	66	12:02:26
32	209.164.26.157	80	202.139.16.45	63450	image/gif	43	12:02:26
33	202.139.16.45	63450	209.164.26.157	80	GET	66	12:02:26
34	209.164.26.157	80	202.139.16.45	63450	image/gif	93	12:02:26
35	202.139.16.45	63450	209.164.26.157	80	GET	66	12:02:26
36	209.164.26.157	80	202.139.16.45	63450	image/gif	43	12:02:26
37	202.139.16.45	63450	209.164.26.157	80	GET	66	12:02:26
38	209.164.26.157	80	202.139.16.45	63450	image/gif	7674	12:02:26
39	209.164.26.157	80	202.139.16.45	63450	image/gif	10119	12:02:26
40	202.139.16.45	63450	209.164.26.157	80	GET	76	12:02:26
41	209.164.26.157	80	202.139.16.45	63450	image/gif	4628	12:02:26
42	209.164.26.157	80	202.139.16.45	63450	image/gif	25482	12:02:26

FIG. 8

Model N1			
Population	Mean	Variance	Weight
	5900	18000	60
Site	Average	Std Dev	Confidence Level
1	2146	3679	45%
2	726	1974	23%
3	1703	4657	37%
4	666	1826	19%
5	1937	4741	39%
6	5149	8433	93%
7	6561	27657	87%

FIG. 9

Model N2			
Threshold			Weight
0.1			20
Site	Text	Graphics	Ratio
1	20617	25938	0.79
2	6402	68319	0.09
3	25821	54455	0.47
4	1701	67037	0.03
5	19019	17252	1.10
6	9869	102919	0.10
7	19446	696989	0.03

FIG. 10

Model N3	
Threshold	Weight
30	20
Site	Word Count
1	0
2	0
3	2
4	0
5	0
6	40
7	37

FIG. 11

Site	Confidence Level
1	27%
2	34%
3	22%
4	31%
5	23%
6	96%
7	92%

FIG. 12